Measurement Instruments in Cross-National Surveys (Version 2.0)

Behr, Dorothée; Braun, Michael; Dorer, Brita

Erstveröffentlichung / Primary Publication
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Behr, D., Braun, M., & Dorer, B. (2016). *Measurement Instruments in Cross-National Surveys (Version 2.0).* (GESIS Survey Guidelines). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.15465/gesis-sg_en_006

gesis Leibniz-Institut für Sozialwissenschaften

GESIS Survey Guidelines

Measurement Instruments in Cross-National Surveys

Dorothée Behr, Michael Braun, & Brita Dorer

December 2016, Version 2.0

Abstract

This contribution deals with measurement instruments – that is, questionnaires and individual questions – used in cross-national survey research. By cross-national surveys we mean studies in which different countries are compared and whose design, implementation, and organisation are shaped by the requirement that comparability be achieved.

We begin by focussing on general questions of comparability in cross-national surveys. With regard to problems that occur generally in all surveys, and to sampling and nonresponse problems in cross-national surveys, we refer the reader to the relevant thematic contributions to the *GESIS Survey Guidelines* and to further literature. We then distinguish between ensuring the comparability of measurement instruments ex ante (e.g., by means of pretests) and ex post (in the context of data analysis or additional surveys). We focus, in particular, on questionnaire translation.

Citation

Behr, D., Braun, M., & Dorer, B. (2016). Measurement instruments in international surveys. *GESIS Survey Guidelines*. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg_en_006

1. General information on measurement instruments in cross-national survey research

This contribution deals with measurement instruments used in cross-national survey research. By cross-national survey research we mean surveys that are conducted in two or more countries and whose design, implementation, and organisation are shaped by the concept of comparability, or equivalence.

Although the focus lies on cross-national surveys, much of what is written here applies also to cross-cultural surveys within one country – for example, surveys that aim to compare different ethnic and linguistic groups within a country.

The structure of the present contribution is based on the premise that both ex-ante and ex-post procedures are needed to create comparable measurement instruments. Comparability must first be established during the design and translation of the questionnaire by means of different processes and measures (Sections 2 and 3). Here, questionnaire design and translation are closely linked – that is, the quality of the former is an essential precondition for the quality of the latter. After collecting the final survey data, their comparability must be tested ex post (Section 4). Section 1 provides a brief general introduction to measurement instruments in cross-national comparative surveys. It addresses the following topics:

1.1 Challenges to comparability of project components

1.2 Measurement error

1.3 Comparability ("equivalence") of the measurement instrument

1.4 Questionnaire types in cross-national survey research

1.1 Challenges to comparability of project components

In cross-national surveys, the problems that must be taken into account are not only those that occur generally in all surveys. Rather, the quality of a cross-national project is measured in terms of the quality – and especially the comparability – of the individual national surveys. If some of the national surveys are affected by errors, the observed cross-national similarities and differences may be merely methodological artefacts. However, even the use of identical procedures in the individual countries may not necessarily guarantee comparability. Indeed, it is sometimes preferable to employ different procedures that are tailored to the respective contexts (e.g., sampling on the basis of different lists or country-specific formulation of response categories). Generally speaking, the following types of errors can be distinguished:

- Do the samples drawn and realised in the various countries represent the population in a comparable way (coverage/sampling/nonresponse error)?
- Do the measurement instruments – that is, the questionnaires and the individual questions after translation – function in the same way everywhere (measurement error)?

Coverage/sampling/nonresponse error is addressed, in particular, in the *GESIS Survey Guidelines* contribution "Nonresponse Bias" (Koch & Blohm, 2016). Measurement error is dealt with in what follows.

1.2 Measurement error

Measurement error is defined as the difference between the observed value and the "true" value. Measurement error has a number of interacting causes that are attributable to the measurement instrument, the interviewers, the respondents, or the survey mode (Biemer et al., 1991).

- Measurement instruments may contribute to measurement error because of their complex design, the wording (and, of course, the inadequate translation) of individual questions, question order, and the response options.
- Interviewers may contribute to measurement error by departing from the question wording, for example, or by incorrectly recording the answers given by respondents. Countries differ in their survey traditions, and thus in their implementation standards, for example in relation to interviewer training.
- Behavioural tendencies on the part of respondents, for example social desirability or acquiescence, are further sources of measurement error. They may be a manifestation of personality traits or be triggered by situational characteristics. There is evidence that the tendency to give socially desirable answers varies across cultures (Johnson & van de Vijver, 2003).
- Moreover, depending on the survey mode, cross-cultural differences in the tendency to answer in a socially desirable way may have a different effect.

In order to ensure comparability between the individual national surveys in a cross-national project, the measurement errors in all participating countries must be minimised as uniformly as possible. Comparing surveys that suffer only from minor measurement errors with those that are characterised by large measurement errors is problematic. For comparison purposes, at least, it would be less problematic to retain the same measurement errors across all countries. This might mean, for example, that sub-optimal questions are asked in all countries and that individual countries refrain from improving questions on their own initiative.

There are a large number of possible measures to mitigate the various above-mentioned problems and their effects.

The *Cross-Cultural Survey Guidelines* (http://ccsg.isr.umich.edu/index.php/chapters/interviewer-recruitment-selection-and-training-chapter) provide a detailed description of how comparability (e.g., in relation to interviewer training and monitoring) can be achieved when administering a survey instrument.

The following sections describe how comparability of the measurement instruments can be ensured and tested.

Readers who would like to engage in depth with the overall challenges associated with cross-national surveys are referred to Braun (2006), Davidov, Schmidt, and Billiet (2011), Harkness et al. (2010), Harkness, van de Vijver, and Mohler (2003), van Deth (2013 [1998]), and van de Vijver and Leung (1997).

1.3 Comparability ("equivalence") of the measurement instrument

The comparability, or equivalence, of the measurement instruments is a prerequisite for any meaningful comparison between individual countries. As a rule, several types of equivalence are distinguished (van de Vijver & Leung, 1997):

- Construct equivalence
- Measurement unit equivalence

- Scalar equivalence

The presence of construct equivalence – that is, the same underlying dimension is measured in the various countries – is a necessary precondition for any substantive comparison between countries. It is also a prerequisite for the other two types of equivalence – measurement unit equivalence and scalar equivalence. To ensure construct equivalence, it may be necessary to use different measurement instruments in different countries. However, the other two types of equivalence then no longer exist. Measurement unit equivalence means that the measurement unit of the scale is identical across countries, but that the origins of the scales are not necessarily identical. Scalar equivalence, on the other hand, means that both the measurement unit and the origins of the scales are identical. This is a prerequisite for the interpretability of the differences in the means across countries, for example.

1.4 Questionnaire types in cross-national survey research

Questionnaire design in a cross-national context may be based on an "ask-the-same-question" (ASQ) approach or an "ask-a-different question" (ADQ) approach.

The underlying assumption on which the ASQ approach is based is that the same questions can be asked in all the countries and cultures participating in a survey. This is typically achieved by first producing a source questionnaire in one project language, usually English, and then translating it into the languages of the participating countries and cultures. ASQ can thus be equated essentially to questionnaire translation.

The ADQ approach is based on the assumption that, although concepts can be jointly defined, their implementation in the form of questions/response categories and/or indicators in the countries and cultures participating in a survey should differ. A country-specific measurement instrument is required when identical questions are not understood – or are understood differently – in different countries, and when this is not due to bad translation but rather to different realities in the various countries. A typical example of a country-specific item is the education question. Even if the wording of the question may be the same across countries, the response categories depend on the education systems of the respective countries. Moreover, when an ADQ approach is applied, the "identity equivalence" method proposed by Przeworski and Teune (1966) can be used (van Deth, 2013): First, identical indicators are determined for all participating countries, and then the respective country-specific indicators are also taken into account. This results in different scales for the individual countries. The functional equivalence of the country-specific items can be tested with the help of the identical items. However, it is often difficult to determine how variants of a question should be worded so that same phenomenon can be captured in all countries. ADQ can be equated to *adaptation*, in the broadest sense of the word.

There are, of course, hybrid forms of questionnaire design. For example, while most of the items in the European Social Survey (ESS) are based on the ASQ approach, some variables, such as education, are collected in a country-specific way (ADQ) and then harmonised according to predefined rules.

The following sections deal with questionnaire design and translation in the context of the ASQ approach.

2. Questionnaire design in cross-national survey research

In what follows, we deal with questionnaire design in the context of the ASQ approach. In addition to the basic substantive and methodological principles that apply to questionnaire design in general, the aim in a cross-national context is to design a questionnaire that is relevant to different cultures and that is, moreover, easily translatable.

Besides the information provided in this section, Harkness, van de Vijver, and Johnson (2003), Harkness, Edwards, et al. (2010), Harkness, Bilgen, et al. (2011), and Smith (2003), for example, offer insights into the design of questionnaires for use in cross-national surveys.

The present section addresses the following topics:

2.1 Cross-national collaboration

2.2 Pretesting

2.3 Problem-oriented coding, translatability assessment, and advance translation

2.1 Cross-national collaboration

Best practice for cross-national surveys requires that the various participating countries should be involved in the design of the questionnaire. Only in this way can it be ensured that the questions developed can be applied cross-nationally and cross-culturally. There are a number of different collaboration models.

In the case of a survey covering, for example, two countries, collaboration is easy to coordinate. However, in multinational studies (e.g., with 20–30 participating countries), coordination and collaboration is more difficult. Viable solutions for large multinational studies can be found in the European Social Survey (ESS) and the International Social Survey Programme (ISSP). In both studies, a group comprising members from different cultures and countries is responsible for developing the questionnaire. Representatives of all participating countries meet at regular intervals to discuss intermediate versions of the questionnaire.

2.2 Pretesting

During the questionnaire design phase, the questionnaire must be tested on the target group. In *quantitative* pretests, the questionnaire intended for use in the main survey is tested under – more or less – realistic conditions. Ideally, quantitative pretesting can be used to test the equivalence of the measurement instruments across participating countries with the help of statistical procedures (see Section 4 below). However, this presupposes that the pretests are implemented with a sufficiently large number of cases in all countries. In practice, and especially in most large-scale cross-national projects, this is not the case. Quantitative pretests often serve merely to investigate (a) comprehension problems with individual questions, (b) the manageability of the instrument as a whole, or (c) the estimation of the time needed to conduct an interview. As a consequence, the quality of the measurement instrument usually has to be tested by secondary researchers in the course of data analysis.

The dominant *qualitative* method is cognitive interviewing, which aims to uncover difficulties that participants have answering the questions (Beatty & Willis, 2007; Willis, 2005). The toolbox of methods has been expanded in recent years to include cross-culturally and cross-nationally comparative cognitive interviews. If these interviews are administered during the questionnaire design phase, problems of non-equivalence can be uncovered and avoided at an early stage.

However, compared to cross-*culturally* comparative cognitive interview studies within one country, *cross-nationally* comparative cognitive interview studies are still very rare (exceptions, e.g.: Fitzgerald et al., 2011, Miller et al., 2011).

More information – including lessons learned and recommendations about how to plan and conduct comparative cognitive interviews – can be found in Farrall et al. (2012), Fitzgerald et al. (2011), Goerman and Caspar (2010), Lee (2012), Miller et al. (2011), and Thrasher et al. (2011).

Because of the considerable effort involved, a situation where large-scale quantitative or qualitative pretests take place in all participating countries almost never occurs in practice. Even in the ESS, qualitative pretests are conducted only in a limited number of participating countries, and, as a rule, a quantitative pretest takes place only in two of the participating countries. Moreover – as is also the case with the ISSP – these pretests are conducted mainly with a view to selecting substantively interesting and methodologically adequate questions rather than as a test of the final measurement instrument to be used in the main survey. Therefore, equivalence testing must be performed additionally ex post – that is, on the basis of the data collected in the main survey (see Section 4).

As a rule, the results of both quantitative and qualitative pretests give rise to – sometimes considerable – revision of the measurement instruments. In principle, the revised versions should also undergo testing. However – usually for time reasons – this, too, is frequently dispensed with in practice.

In what follows, we present a special type of testing that can also be conducted post hoc – namely, the application of probing techniques in cross-national web surveys. To analyse an ISSP item about attitudes to civil disobedience, Behr et al. (2014) used both comprehension probes ("What ideas do you associate with the phrase 'civil disobedience'? Please give examples.") and category-selection probes ("Please explain why you selected [category]?"), which were incorporated into cross-national web surveys after the ISSP item in question. With the help of the answers entered by the respondents, the authors could demonstrate that the low level of support for civil disobedience in countries such as the USA and Canada, compared to countries like Germany and Spain, was due essentially to two reasons: First, civil disobedience was more strongly associated with violence in the first-named group of countries than in the second group. This is a methodological artefact – that is, respondents answered essentially different questions. Second, trust in politicians was (even) lower in the second group than in the first. This is a substantive result. Hence, a methodological artefact and real differences are effectively confounded in the data.

Moreover, Braun, Behr, and Kaczmirek (2013) demonstrated in the case of ISSP items relating to attitudes towards migrants that – despite the fact that respondents from different countries thought mostly of comparable groups who reflected the actual composition of the migrant population in the country in question – characteristic biases nonetheless occurred. The authors used "specific probes" ("What groups were you thinking of when you were answering the question?"), which were implemented after the aforementioned ISSP items. The migrant groups who came to mind when the respondents were answering the questions could not have been investigated with standard statistical procedures alone.

2.3 Problem-oriented coding, translatability assessment, and advance translation

Coding schemes for the systematic appraisal of items – also known as *question appraisal systems* – can also be incorporated into the questionnaire design process. The aim of these coding schemes is to systematically test new items for certain quality dimensions. Dean et al. (2007) proposed a coding scheme specifically tailored to the needs of cross-cultural surveys in the USA.

Furthermore, Aquadro et al. (2012) developed a type of *translatability assessment* for medical questionnaires. Here, too, a source question is systematically tested for its intercultural applicability and translatability.

Advance translation pursues similar goals and objectives. This method entails producing an advance translation during the questionnaire design phase in order to make the source questionnaire easier to translate and thereby help to minimise subsequent translation problems and errors. The objective is to formulate the – usually English-language – source questionnaire in such a way that (a) it contains as few elements as possible that will cause problems during subsequent translation into different languages, and (b) the fact that the questionnaire is to be subsequently administered in the target languages is also taken into account during the questionnaire design phase. The advance translation can help to identify elements of the questionnaire that are problematic with regard to (a) and (b) above. When producing an advance translation, teams comprising experienced questionnaire translators and survey methodologists should jointly translate a provisional version of the source questionnaire and specifically comment on those parts of the text where they encountered problems during the translation process. Moreover, team members should be encouraged to make any type of comments that come to mind with regard to the subsequent use of the questionnaire in the various target cultures. Since 2009, advance translation has been an integral part of the design of the English-language source questionnaire of the ESS, for example (Dorer, 2011).

The Survey Quality Prediction (SQP) programme focuses on evaluating formal questionnaire design criteria. Detailed information can be found on the programme website, http://sqp.upf.edu/.

3. Questionnaire translation

The quality of questionnaire translations is of decisive importance for the quality of research results.

Section 3 addresses the following topics:

- 3.1 What should be taken into account when translating questionnaires?
- 3.2 How should the translation process be organised?
- 3.3 Must a translated questionnaire be tested separately?
- 3.4 Where can I find translators?
- 3.5 What information should be made available to translators?
- 3.6 I need translations not only for another country but also for a particular migrant group in that country. What should I take into account in this case?
- 3.7 Are questionnaire translations always one and the same thing? Are there different types of questionnaire translations?
- 3.8 What types of question/questionnaire adaptations are there?
- 3.9 What should be taken into account in the case of project proposals that provide for a cross-national or cross-cultural survey to be conducted?

3.1 What should be taken into account when translating questionnaires?

First, the quality requirements that a successful questionnaire translation must meet are the same as those that apply to any other kind of translation: It must be substantively and linguistically correct and complete. Moreover, a translated questionnaire, as a survey instrument, must work as intended in the target language. Hence, it is particularly important that the *meaning* is expressed in a generally

understandable way in the target language. Therefore, the stimulus in the source questionnaire should correspond to that in the target questionnaire.

Because a cross-section of the total population is usually surveyed in social science surveys, it is particularly important that the language used should be as simple and easily understandable as possible. Complicated sentence constructions should be avoided, as should rare words and, of course, foreign words. Questions and predefined response categories should also be clearly and unambiguously worded in the target language.

In case of doubt, the survey practices that are customary in the target country should serve as a guide – for example, with regard to the way in which the respondents should be addressed (e.g., politeness) or the way grammatical gender should be handled in survey questions. In Germany, for example, one should decide whether to exclusively use the masculine form to cover males and females alike, or whether the masculine and feminine forms should always be explicitly used side by side (e.g., *Wie stolz sind Sie, ein/e Deutsche/r zu sein?*).

Care should also be taken to ensure consistency within the questionnaire. This applies not only to the consistent use of individual key words and terms within the questionnaire but also to the consistent use of translated words and terms across different survey rounds or waves. This is especially relevant in the case of panel surveys and repeated cross-sectional surveys, where, in the interests of comparability over the years, it is important that certain terms and wording be used consistently over time.

A detailed list and discussion of different points to which attention should be paid when translating questionnaires can be found in the second part of the *ESS Translation Guidelines* (European Social Survey, 2014).

3.2 How should the translation process be organised?

The so-called team approach, or committee approach, to questionnaire translation has established itself as best practice in social science survey research (see, for example, European Social Survey, 2014; Harkness 2003; Harkness, 2010). This approach is characterised, first, by the fact that several people with different skills and disciplinary expertise participate in the translation, and, second, that a multistage process is used. The advantages of this approach are as follows: The fact that one person is not relied on to produce the translation minimises the risk that the end-product will be too strongly shaped by the personal style of just one individual. Moreover, even the most experienced translators make mistakes, and it often happens that even very well-trained translators understand one and the same word differently. By including different people in the process, these sources of error are reduced. The fact that the translation passes through several procedural steps is an additional safety mechanism, so that the probability of incorrect translation or an excessively personal translation style can be minimised until the final translation is completed.

The team approach should be supplemented by different pretest procedures and careful process documentation. The entire process is also known as "TRAPD". Originally developed by Janet Harkness, it is described, for example, in Harkness (2003) and in the *ESS Translation Guidelines* (European Social Survey, 2014). TRAPD is an acronym formed by combining the initial letters of the five procedural steps in this process:

- "T" is for translation, the first step, which entails producing two independently executed translations. It is recommended that at least one of the translators should be a trained and/or professional translator or at least have a degree in linguistics.
- "R" is for review, the subsequent session moderated by a reviewer, during which the two independently produced translations are discussed and, if possible, a solution is found for each questionnaire item.

- "A" is for adjudication, the phase in which any uncertainties that persist after the review step are clarified with the help of an experienced survey expert, the adjudicator. People who are chosen to act as reviewers and adjudicators should have relevant knowledge of social science and/or of survey methodology (on the selection of suitable people, see Section 3.4).
- "P" is for the pretest in which the translations of all the questionnaire items are tested on a small sample of the target population in order to estimate the extent to which they are understood in the manner intended (see Section 3.3).
- "D" is for documentation, the last step in the entire process. The aim here is to make information about how the final versions of the translations came about available to interested parties, if possible both within and outside the project in question.

There are, of course, other methods for producing a questionnaire translation. In what follows, we will focus, in particular, on one of these methods, namely back translation. This method is still in use, despite the fact that its weaknesses have often been pointed out (see, for example, Harkness, 2003; Schoua-Glusberg & Villar 2013). It involves, first, having one person produce a translation from the source language, A, into the target language, B, and then having another person translate this translation back into language A. The source text in language A is then compared with the back translation into language A. If the two versions differ, the translation is checked once again and, if necessary, revised. However, from a translation sciences perspective, this method is questionable. Good translations are often characterised precisely by a certain degree of freedom. It is therefore very unlikely that the back translation of a good – that is, a relatively free – translation will result in the same text as the source text. This means that good translations may be wrongly deemed to be problematic. On the other hand, bad translations may go unnoticed if, for example, both translators (A-B and B-A) who participate in the process translate badly. Janet Harkness (Harkness, 2003) chose a good example to illustrate this problem, namely the German idiom "*das Leben in vollen Zügen genießen*" (to live life to the full). If the translator translates it unidiomatically – that is, badly – as "to enjoy life in full trains," and the literal back translation into German yields exactly the same wording, the fact that it is a bad translation may go unnoticed. If, on the other hand, the TRAPD approach is applied, it is very likely that such an inadequate translation would be noticed and pointed out by some of the people involved in the individual procedural steps. This, of course, presupposes that those who participate in a TRAPD process are qualified for the task. Even if a back translation is provided for in the respective survey specifications, it is not advisable to regard it as the sole – or as an adequate – quality assurance method.

3.3 Must a translated questionnaire be tested separately?

A translated questionnaire must be tested as carefully on the target group as a questionnaire developed in the target language. On the one hand, qualitative test methods can be used – the *GESIS Survey Guidelines* contribution "Cognitive Pretesting" (Lenzner, Neuert, & Otto, 2016) provides guidance on the concrete implementation of cognitive interviews. On the other hand, quantitative methods can be employed, which allow statistical analyses to be conducted. These quantitative approaches include the split-ballot method, which involves administering different versions of some questions to a randomly split sample and analysing the responses (Fowler, Jr., 2004; Krosnick, 2011). The quantitative pretest can be supplemented with a so-called debriefing, which involves gathering additional information about the course of the interview after it has been conducted (Martin, 2004).

However, if the aim of translating the questionnaire is to achieve comparability with the source text, a number of particularities apply to the testing of translated questionnaires:

1. If the pretest uncovers problems with the specific wording of the translation, they can often be solved by rewording the translation. For example, a term that was not understood, or that was understood incorrectly, can be replaced with a more appropriate term or can be paraphrased.
2. If the test uncovers problems with the translated questionnaire that are not related to the actual wording, but rather are of a general nature (e.g., the recall period is too long), these problems cannot be solved if the source text has already been finalised and can no longer be modified. Ideally, therefore, in order to be able to identify and subsequently solve translation, cultural, and general problems at an early stage, pretests should be conducted while the source questionnaire for a cross-national survey is being developed (see Section 2.2).

Statistical testing of translated questionnaires – Discipline-specific particularities

In psychology and health research, the psychometric testing of translated questionnaires for (a) reliability (see also the *GESIS Survey Guidelines* contribution "Reliability," Danner, 2016), (b) validity, (c) sensitivity/specificity (in health research), and/or (d) equivalence with the source questionnaire is part of the overall concept for the production of a questionnaire translation. These key data are frequently published in research articles.

Although translated questionnaires also undergo statistical testing in the social sciences in the context of quantitative pretesting, it is left up to the secondary researchers, in particular, to test the quality (e.g., the equivalence) of the data from cross-national surveys such as the ESS or the ISSP before they use them. However, there have recently been calls for statistical quality testing in the context of questionnaire design (and translation) in social survey research. In response to these calls, Rammstedt et al. (2015) proposed quality standards as a general guideline.

The relevant literature in the various (sub-)disciplines should be consulted with regard to recommended statistical procedures. Keywords for literature searches are *equivalence*, *validation*, *psychometric testing*, and *adaptation*. Section 4 below also provides some insights.

3.4 Where can I find translators?

Finding suitable translators is quite a challenge because the area of questionnaire translation is a niche in which very few translators have specialised. Recommendations from other surveys, from national translators associations – in Germany, for example, the *Bundesverband der Dolmetscher und Übersetzer e.V.* (BDÜ) and the *Fachverband der Berufsübersetzer und Berufsdolmetscher e.V.* (ATICOM) – or from universities specialising in translation can be first points of contact when searching for suitable personnel.

Overall, when selecting translators, it should be ensured that they

1. translate into their mother tongue;
2. have a very good knowledge of the source language;
3. have translation competence (not necessarily proven by a degree in translation); however, translation competence consists of much more than understanding a foreign language more or less (see EMT, 2009);
4. are sensitive to different registers and have a good feel for language in general in order to be able to translate in a manner that is suitable for the target group;
5. are willing to actively participate in translation discussions if the above-mentioned TRAPD approach is applied (this includes a willingness to constructively and critically appraise their own translations and those of third parties);

6. are, ideally, familiar with the particularities of questionnaire translation;
7. have, ideally, a knowledge of survey design and the operationalisation of concepts;
8. have, ideally, expertise in the subject area in question.

Especially with regard to the last three points, a lack of the necessary knowledge on the part of the translators can be offset through appropriate briefing and/or training. Moreover, the team approach described above ensures that the necessary expertise can be contributed by other protagonists in the translation process – for example, social scientists or psychologists with expertise in the subject area in question.

Here, the reader is referred to the *Cross-Cultural Survey Guidelines*, and especially to the suggestions for the recruitment of translators: http://ccsg.isr.umich.edu/index.php/chapters/translation-chapter/finding-selecting-and-briefing-translation-team-members

3.5 What information should be made available to translators?

The amount and type of information made available to the translators depends on various factors. Is a particular project or survey demanding in terms of technical terminology, or must a specific target audience be taken into account when choosing the register? Are the translators already experienced in translating questionnaires, or must they first be briefed in specific demands of this text genre?

The following points can serve as a guideline for the information that should be made available to the translators:

1. Background information on the survey in question, for example about the objectives and content of the survey or the design of the questionnaire
2. Target audience, for example in terms of age, education, nationalities, specific population- or occupational groups
3. Survey mode: Will the survey be administered orally or in the form of a paper-and-pencil or web-based self-administered questionnaire?
4. Will this questionnaire be translated only into one language or into other languages as well?
5. Are translations to be integrated into existing translations – that is, must existing translations be used? This is of particular relevance in the case of repeated cross-sectional surveys and panel surveys, where, as a rule, only some of the questions have to be newly translated. However, these newly translated questions must blend in seamlessly with the repeated questions, for example with regard to the type of question wording and the wording of the response categories.
6. Is the source text final, or may changes still take place, for example because of questions that arise during the translation process?
7. Whom should the translators contact if they have any questions?
8. What method is used to review or ensure the quality of the translation?

Moreover, the translators should always be informed of the points mentioned in Section 3.1 above relating to the general guidelines that apply to questionnaire translations (e.g., linguistic correctness, completeness, comprehensibility, observance of norms of politeness, consistency).

3.6 I need translations not only for another country but also for specific migrant groups in that country. What should I take into account in this case?

In principle, the rules and requirements that apply in the case of a questionnaire translation for a migrant group in a particular country are the same as those discussed in the present guidelines. Hardly any research has been done to date on what distinguishes translations into migrant languages from translations into the languages of other countries. If at all, this research has been conducted in the U.S. context. However, practical experience in the USA has revealed a number of aspects, which can be summarised as follows:

1. What country or countries is the target group from? This should be taken into account when producing the translation. If, for example, a questionnaire is provided in Spanish, the countries of origin of the target population(s) (e.g., different Spanish-speaking countries with possible linguistic differences) should be taken into account when formulating questionnaire items (Martinez, Marín, & Schoua-Glusberg, 2006; Alegria et al., 2004).
2. It sometimes makes sense to forgo translating some names – for example, those of certain state bodies or programmes – because migrants are more familiar with the designation in the source language than in their mother tongue. In connection with the translation of the Austrian *Mikrozensus*, Stadler (2010) cited the term *Arbeitsmarktservice*, which is the Austrian designation for economic stimulus programmes. When it was translated into the migrant languages in which the survey was made available, the migrants did not understand it.
3. If questions relate to Germany, for example, the translation should be formulated in such a way that it is clear that Germany is the frame of reference. In other words, the translation should not suggest that the frame of reference is the migrant's country of origin. The use of the source-language designation for a concept (see point 2 above) may also be helpful in this case.
4. When producing translations, attention should be paid to the fact that migrants and natives may have a different perspective of the target country. What migrants consider to be *foreign*, *another country*, *a foreign country*, or *abroad* is not necessarily the same as what natives understand by it. Therefore, when producing a translation, the respective frames of reference should be critically examined. By way of example, one can cite here the Chinese and Korean translations of a question fielded in the American Community Survey conducted by the U.S. Census Bureau: "Where was this person born?" One of the response categories read: "Outside the United States – Print name of foreign country, or Puerto Rico, Guam, etc." In both the Chinese and Korean translations, "foreign country" was translated only as "country" in order to avoid confusing respondents, who were supposed to report their home country (Sha, Park, & Pan, 2012).
5. Questions may be understood differently against a different cultural background. The first version of the translation into Korean of the (U.S. Census Bureau) question "In what year did this person[1] last get married?" caused some problems. Many Korean participants in a cognitive pretest regarded marriage as a once in a lifetime event. Against this background, the translation was interpreted as meaning "When did the person get married recently?", with the result that those who had been married for a long time thought that the question did not apply to them. The problem was solved by dropping the word "last" in the question and including an additional instruction for respondents (Sha et al., 2012).
6. Questions in national surveys may be considered by migrants to be taboo subjects and cannot therefore be asked (Sauer, 2008).
7. A further problem arises when migrant groups do not – or no longer – speak the language of their country of origin in the way that it is spoken there. This may be due, for example, to the

[1] "This person" means the respondent.

fact that emigrants from the country in question belonged mainly to a specific group, or that elements of the language of the host country have been assimilated. This should be taken into account when searching for suitable translators.

3.7 Are questionnaire translations always one and the same thing? Are there different types of questionnaire translations?

The previous sections dealt with questionnaire translation in the context of intentionally oriented cross-national surveys. However, there are other types of questionnaire translations that call for (cultural) adaptations, to a greater or lesser extent. They are presented below.

When the translation of a questionnaire is pending, the following key questions can help to obtain an initial overview of the general demands and constraints to which it is subject. For questionnaire translations are not always one and the same thing, as will become clear in what follows.

The key questions are:

1. Was the source questionnaire – that is, the questionnaire to be translated – originally designed for a particular culture or country?

   OR

   Was the source questionnaire designed with a view to cross-national application?

2. Is the aim of the questionnaire translation to establish comparability with the source questionnaire?

   OR

   Does the source questionnaire serve merely as a source of ideas for a new questionnaire in another language?

These questions yield the following table:

Table 1: Different types of questionnaire translations

	Comparability with source questionnaire planned	Comparability with source questionnaire <u>not</u> planned
Source questionnaire designed for a particular culture/country	I: Adaptations probably necessary; however, they are constrained by the comparability requirement	II: Adaptations probably necessary, no constraints
Questionnaire designed with a view to cross-national application	III: Adaptations possibly necessary; permissible adaptations often expressly identified as such; in cross-national surveys, further adaptations are usually subject to authorisation	IV: Adaptations possibly necessary, no restrictions

The four table cells will now be discussed with a view to determining where the questionnaire translation might be located on the continuum between translation and adaptation (Behr, 2013). In this connection, *adaptation* stands for modifications that go beyond those "typically" made during translation and that are necessary in order to adapt the questionnaire to the conditions prevailing in another cultural area. On specific adaptations, see Section 3.8 below.

Situation I

If a questionnaire that was developed in the U.S. context, for example, is to be used as a basis for a German questionnaire, and if the aim is to achieve the greatest possible comparability, the following applies: On the one hand, adaptations are necessary in order to detach the questionnaire from its American anchorage and to adapt it to living conditions in Germany. On the other hand, however, modifications must be restricted to a minimum because, otherwise, the comparability of the data cannot be guaranteed.

An adaptation would be necessary in the case of the following item, for example: "Do you have difficulty walking several blocks?" City blocks are a central element of American urban design; depending on the measurement intention, "several blocks" could be replaced in German with a distance in metres or an equivalent paraphrase (Harkness, Edwards, & Villar, 2010).

Situation II

If a questionnaire that was originally developed for a U.S. context, for example, is now to be used as a basis for a German questionnaire, and if data comparability is not the aim, the translation is, in principle, free of constraints. Changes may be necessary for cultural reasons, or may simply be aimed at optimising the measurement instrument for your own research purposes – for example, by changing a 7-point scale into a 5-point scale or by adding or deleting introductions to topics or whole questions, etc.

Situation III

If a source questionnaire is newly designed for a cross-national comparative survey, it is, ideally, formulated in such a way that either it is translatable in the participating countries or permissible adaptations (e.g., the addition of country-specific examples or response categories) are expressly identified as such in advance. As a rule, further-reaching adaptations are subject to authorisation by a central coordinating body and are restricted to a minimum. Situation III corresponds, for example, to the situation in the ESS and the ISSP. This situation forms the basis for the description provided in Sections 1 and 2 above.

Situation IV

If a questionnaire is to serve merely as a source of ideas for a questionnaire that is to be newly developed, one is free to make the desired changes to the questionnaire during translation. As the source questionnaire was developed with a view to cross-national application, a number of questions may be translatable and do not have to be adapted.

In addition to the conditions outlined above, what can still be translated or what must be adapted in a specific case, and what adaptations are permissible, depends on the following factors:

Topic

What is the topic of the questionnaire? Typically, general questions about satisfaction with different areas of life are more translatable than questions that address specific structures of the health system, for example.

Linguistic and cultural distance

How great is the distance between the source question and culture and the target question and culture? The greater the respective distances, the more likely it is that adaptation will be necessary.

Survey specifications

What requirements are laid down in the survey specifications? This applies, in particular, to large-scale cross-national surveys, which usually specify the degree of freedom that the participating countries have when transferring a questionnaire into another language.

Legal constraints

Is the questionnaire subject to copyright? This applies more to questions from the field of psychology or health research than to questions from social science disciplines. In the case of material protected by copyright, translations and adaptations are subject to authorisation by the respective right holders.

In addition to the above-mentioned questionnaire translations, which aim to produce a functioning measurement instrument for another cultural area, there is a further type of questionnaire translation, namely "documentary" questionnaire translation. The German General Social Survey (ALLBUS), for example, is translated into English with the aim of giving scientists who do not speak German an insight into the survey. It is explicitly pointed out that the translated questions would have to be adapted for use in an English-speaking country.

3.8 What types of question/questionnaire adaptations are there?

The term *adaptation* is understood differently not only by different scientists but also across different disciplines (Behr, 2013).

On the one hand, *adaptation* is used as a superordinate term for the overall process of transferring a measurement instrument from one language/culture to another. It is employed in this way especially in health research and cross-cultural comparative psychological research to indicate that a (semantically comparable) translation is sometimes not sufficient and that modifications have to be made at different levels. Moreover, in its superordinate form, this term also serves to draw attention to the necessity of psychometrically testing a new measurement instrument.

On the other hand, *adaptation* is used to designate various specific modifications at the item level. In this context – to put it simply – one can distinguish between adaptations in the area of culture, measurement, and language. Adaptations in the area of language are often also a prerequisite for a perfectly normal, good *translation*.

Adaptations in the area of culture

This covers modifications that are determined by differences between systems (in the broadest sense of the word) in the individual countries and cultures. Specific references in the questionnaire to the political system, the education system, the health system, etc. often require adaptation. This applies also to references to units of measurement (e.g., miles/metres). Moreover, modifications of the layout are often necessary to take account of different naming, date-format, and address-format conventions, for example.

In addition to the aforementioned modifications to the questionnaire that are determined more by factual differences, there are modifications that are determined by different values, customs, or habits.

The question "Have you ever worn a campaign badge/sticker" should not be asked in countries in which badges/stickers are not part of the inventory of symbols of political participation. The question "How difficult was it to sit without losing your balance?" measures the extent to which respondents can maintain their balance while sitting on a chair. When translating the question into Japanese, the still common Japanese tradition of sitting on the floor should be taken into account. The intended measurement can be ensured by making the question more specific: "How difficult was it to sit on a chair without losing your balance?" (Breugelmans, 2009).

Other thematic areas that often require adaptation are family-related topics, religious topics, leisure activities, food, and medicines.

Adaptations in the area of measurement

This covers modifications that are determined by different experience with measurement instruments and surveys. When transferring a questionnaire into another language and culture, instructions/explanations/images relating to scale labels, etc. may have to be supplemented if the new target group has little or no experience with measurement instruments.

Moreover, scales may have to be adapted in order to take account of the norms of social discourse in the target culture. One example that is sometimes cited in this regard is the adaptation of the agree-disagree scale in Japanese by softening the English-language endpoint labels *strongly agree* and *strongly disagree* (Smith, 2004).

Adaptations in the area of language

This covers changes that are necessitated simply by differences between linguistic systems – for example, changes in word order, parts of speech, gender specifications, etc. Although these changes are referred to in the literature as "language-driven adaptations" or "linguistics-driven adaptations" (e.g., Harkness, 2008), they should in fact be perceived as key characteristics of a good *translation*. Indeed, translation in itself should not be understood as a rigid word-for-word transfer but rather as the transfer of meaning, taking into account the linguistic and idiomatic demands of the target language.

During translation, the questionnaire must also be adapted to the norms of discourse of the target culture. For example, account should be taken of different conventions of politeness when addressing people or formulating questions/instructions. These changes are also an essential prerequisite for a good *translation*, as they cause the text to be perceived as an original work rather than a translation.

If the new target group is a population group that has a lower level of education than the population group to whom the original questionnaire was addressed, the language of the questionnaire may have to be simplified. Irrespective of educational differences, terms that are not understood by the target group should be paraphrased or explained. Here, the boundary is undoubtedly blurred between what some people classify as "still translation" and what others classify as "already adaptation".

3.9 What should be taken into account in the case of project proposals that provide for a cross-national or cross-cultural survey to be conducted?

In the past, at least, it was not unusual in substantive research projects for questionnaire translation to become an issue only when it was actually pending; time and money for questionnaire translation were rarely budgeted for.

Against the background of the far-reaching consequences that a bad translation can have for the research results, account should be taken of the multidimensionality of translation and the connection between good questionnaire translation and good questionnaire design. If a proposal for a new project provides for a cross-national or cross-cultural survey to be conducted, the following applies:

1. The cross-national/cross-cultural perspective should be taken into account when designing the source text by involving (groups of) people from other countries or cultures and by applying different qualitative and quantitative test methods.
2. Questionnaire translation calls for the involvement of different (groups of) people and for a combination of different quality assurance methods.
3. A cross-national or cross-cultural survey has consequences for the time schedule and the personnel- and financial planning of a research project.

4. Testing measurement instruments ex post

There are a number of quantitative procedures for testing equivalence ex post. By far the most frequently applied procedure is confirmatory factor analysis (Brown, 2006; Vandenberg & Lance, 2000). In the case of cross-cultural comparisons, multi-group confirmatory factor analysis is recommended. This method involves testing whether a measurement instrument that has proved its worth in one country also works well in the other countries. Comparability can exist at different levels that correspond to the equivalence criteria discussed in Section 1.3 above: Configural invariance holds true if all items in the various countries belong to the same latent dimension; metric invariance requires that the items have equal factor loadings; scalar invariance additionally presupposes that all intercepts are the same across countries. The studies on basic human values conducted by Davidov (2008) and Davidov, Schmidt, and Schwartz (2008) are an example of such group comparisons by means of confirmatory factor analysis.

Multilevel analysis is a statistical method that explicitly takes account of the existence of different levels – the respondent level and the country level (Hox, 2010; Snijders & Bosker, 2011). Multilevel analysis is especially useful when the number of countries is large (20 is often considered to be the minimum) and when country-level variables are to be explicitly included to explain equivalence problems. Multilevel analysis can also be combined with confirmatory factor analysis (multilevel confirmatory factor analysis). However, this calls for a considerably greater number of cases at the country level.

One method that is used in psychology, in particular, is item response theory, which analyses the relations between the individual items and the underlying dimension (Embretson & Reise 2000, van de Vijver & Leung, 1997). Item response theory calls for a large number of items, which limits its applicability in many cross-national social science surveys.

There are a number of other methods for testing equivalence (and, thus, comparability across countries), which, while not necessarily offering the same degree of precision, are sometimes much easier to use. For example, instead of confirmatory factor analysis, exploratory factor analysis can be performed (van de Vijver & Leung, 1997). (Multiple) correspondence analysis (Greenacre & Blasius 2006) is essentially a principal component analysis method for categorical data. Blasius and Thiessen (2006) applied correspondence analysis to the ISSP gender role items. Multidimensional scaling (Borg & Groenen 2005) graphically represents the intercorrelations between the items. The data must be at least ordinal. The criterion for comparability is whether the items in the individual countries are located in

the appropriate regions. The boundaries between the regions are derived by theoretical reasoning. Braun & Johnson (2011) applied a variety of – complex and less complex – procedures to the same substantive problem and demonstrated that, in many cases, they led to the same conclusions.

One disadvantage of all quantitative procedures – with the possible exception of multilevel analysis when country-level variables are included that can (potentially) explain the inequivalences – is that they identify the presence of problems but cannot explain the actual causes. Hence, the use of probing techniques is recommended, as in the case of ensuring equivalence ex ante (i.e., in the context of pretests). As cognitive interviews, and especially cross-nationally comparative cognitive interviews, are extremely costly and time-consuming to carry out, the implementation of additional web-based studies would appear to be a possible alternative. In this case, following Schuman's proposal (1966) for "random probes," probing questions can be incorporated into a normal web questionnaire. This approach was applied in the studies conducted by Behr et al. (2014) and Braun, Behr & Kaczmirek (2013), which were described in Section 2.2 above.

References

Alegría, M., Vila, D., Woo, M., Canino, G., Takeuchi, D., Vera, M., ... Shrout, P. (2004). Cultural relevance and equivalence in the NLAAS Instrument: Integrating etic and emic in the development of cross-cultural measures for a psychiatric epidemiology and services study of Latinos. *International Journal of Methods in Psychiatric Research, 13*(4), 270-288.

Behr, D. (2013). On the different uses and users of the term adaptation. Paper presented at the AAPOR conference, 16.05.-19.05.2013. Boston, MA.

Behr, D., Braun, M., Kaczmirek, L., & Bandilla, W. (2014). Item comparability in cross-national surveys: Results from asking probing questions in cross-national web surveys about attitudes towards civil disobedience. *Quality & Quantity, 48*, 127-148.

Biemer, P.P., Groves, R.M., Lyberg, L., Mathiowetz, N.A., & Sudman, S. (Eds.) (1991). *Measurement errors in surveys*. New York: Wiley.

Blasius, J., & Thiessen, V. (2006). Assessing data quality and construct comparability in cross-national surveys. *European Sociological Review, 22*, 229-242.

Borg, I., & Groenen, P.J.F. (2005). *Modern multidimensional scaling* (2nd ed.). New York: Springer.

Braun, M. (2006). *Funktionale Äquivalenz in interkulturell vergleichenden Umfragen. Mythos und Realität*. Mannheim: ZUMA.

Braun, M., Behr, D., & Kaczmirek, L. (2013). Assessing cross-national equivalence of measures of xenophobia: Evidence from probing in web surveys. *International Journal of Public Opinion Research, 25*, 383 395. doi:10.1093/ijpor/eds034.

Braun, M., & Johnson, T.P. (2010). An illustrative review of techniques for detecting inequivalences. In Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.-E., & Smith, T. (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 375 393). Hoboken, NJ: Wiley.

Breugelmans, R. (2009). Dangers in using translated medical questionnaires. The importance of conceptual equivalence across languages and cultures in patient-reported outcome measures. *Chest, 136*, 1175-1177.

Brislin, R. W. (1986). The wording and translation of research instruments. In W.J. Lonner & J. W. Berry (Eds.), *Field Methods in Cross-Cultural Research* (pp. 137-164). Beverly Hills: Sage.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Danner, D. (2016). Reliability – The precision of a measurement. *GESIS Survey Guidelines*. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg_en_011

Davidov, E. (2008). A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey. *Survey Research Methods, 2*, 33-46.

Davidov, E., Schmidt, P., & Billiet, J. (2011). *Cross-cultural analysis: Methods and applications*. New York: Routledge.

Davidov, E., Schmidt, P., & Schwartz, S.H. (2008). Bringing values back in: The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly, 72*, 420-445.

Dean, E., Caspar, R., McAvinchey, G., Reed, L., & Quiroz, R. (2007). Developing a low-cost technique for parallel cross-cultural instrument development: The question appraisal system (QAS-04). *International Journal of Social Research Methodology, 10*, 227-241.

Dept, S. (2013). *Translatability assessment of draft questionnaire items*. Paper presented at the 5th Conference of the European Survey Research Association (ESRA), Ljubljana. http://www.europeansurveyresearch.org/conference/programme?sess=38&tday=1

Dorer, B. (2011). *Advance translation in the 5th round of the European Social Survey* (ESS). FORS Working Paper Series, paper 2011-4. Lausanne: FORS

EMT expert group (2009). *Competences for professional translators, experts in multilingual and multimedia communication* http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf

European Social Survey (2014). *ESS Round 7 translation guidelines*. London: ESS ERIC Headquarters, Centre for Comparative Social Surveys, City University London.

Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah NJ: Lawrence Erlbaum.

Farrall, S., Priede, C., Ruuskanen, E., Jokinen, A., Galev, T., Arcai, M., & Maffei, S. (2012). Using cognitive interviews to refine translated survey questions: An example from a cross-national crime survey. *International Journal of Social Research Methodology, 15*, 467-483.

Fitzgerald, R., Widdop, S., Gray, M., Collins, D. (2011) Identifying sources of error in cross-national questionnaires: Application of an error source typology to cognitive interview data. *Journal of Official Statistics, 27*, 569-599.

Fowler, Jr., F. J. (2004). The case for more split-sample experiments in developing survey instruments. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 173-171). Hoboken, NJ: Wiley.

Goerman, P.L., & Caspar, R. A. (2010). A preferred approach for the cognitive testing of translated materials: Testing the source version as a basis for comparison. International *Journal of Social Research Methodology, 13*, 303-316.

Greenacre, M. J., & Blasius, J.. (Eds.) (2006). *Multiple correspondence analysis and related methods*. Boca Raton, FL: Chapman & Hall.

Harkness, J. (2003). Questionnaire translation. In J. Harkness, F. van de Vijver, & P. Mohler (Eds). *Cross-cultural survey methods* (pp. 35-56). Hoboken, NJ: Wiley.

Harkness, J. A. (2008). Comparative survey research: Goal and challenges. In E. de Leeuw, J. J. Hox & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 56-77). New York: Lawrence Erlbaum Associates.

Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.-E., & Smith, T. (Eds.) (2010), *Survey methods in multinational, multiregional, and multicultural Contexts* (pp. 375 393). Hoboken, NJ: Wiley.

Harkness, J., van de Vijver, F.J.R., & Johnson, T.P. (2003). Questionnaire design in comparative research. In Harkness, J., van de Vijver, F.J.R., & Mohler, P.P. (Eds.), *Cross-cultural survey methods* (pp. 19-34). Hoboken, NJ: Wiley.

Harkness, J.A., Edwards, B., Hansen, S.E., Miller, D.R., & Villar, A. (2010). Designing questionnaires for multipopulation research. In Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.-E., & Smith, T. (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (33-58). Hoboken, NJ: Wiley.

Harkness, J., Bilgen, I., Córdova Cazar, A., Hu, M., Huang, L., Lee, S., Mingnan, L., Miller, D., Stange, M., Villar, A., & Yan Ti. (2016). Cross-cultural survey guidelines. Questionnaire Design. URL: http://ccsg.isr.umich.edu/index.php/chapters/questionnaire-design-chapter

Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 117-140). Hoboken, NJ: Wiley.

Harkness, J., van de Vijver, F.J.R., & Mohler, P.P. (Eds.) (2003), *Cross-cultural survey methods*. Hoboken, NJ: Wiley.

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (2nd ed.). New York: Routledge.

Johnson, T.P., & van de Vijver, F.J.R. (2003), Social desirability in cross-cultural research. In Harkness, J., van de Vijver, F.J.R., & Mohler, P.P. (Eds.) (2003), *Cross-cultural survey methods* (pp. 195-204) Hoboken NJ: Wiley.

Koch, A., & Blohm, M. (2016). Nonresponse bias. Mannheim, GESIS – Leibniz Institute for the Social Sciences (*GESIS Survey Guidelines*). doi: 10.15465/gesis-sg_en_004

Krosnick, J. A. (2011). Experiments for evaluating survey questions. In J. Madans, K. Miller, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: contributing to the science of data collection* (pp. 215-238). Hoboken, NJ: Wiley.

Lee, J. (2012). Conducting cognitive interviews in cross-national settings. *Assessment*, published online before print on 11 February, 2012. doi: 10.1177/1073191112436671

Lenzner, T., Neuert, C., & Otto, W. (2016). Cognitive Pretesting. *GESIS Survey Guidelines*. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg_en_010

Martin, E. (2004). Vignettes and respondent debriefing for questionnaire design and evaluation. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 149-188). Hoboken, NJ: Wiley.

Martinez, M., Marín V., & Schoua-Glusberg, A. (2006). Translating from English to Spanish: The 2002 national survey of family growth. *Hispanic Journal of Behavioral Sciences, 28*(4), 531-545.

Miller, K., Fitzgerald, R., Padilla, J.-L., Willson, S., Widdop, S., Caspar, R., Dimov, M., Grey, M., Nunes, C. Prüfer, P., Schöbi, N., & Schoua-Glusberg, A. (2011). Design and analysis of cognitive interviews for comparative multinational testing. *Field Methods, 23*, 379-396.

Rammstedt, B., Beierlein, C., Brähler, E., Eid, M., Hartig, J., Kersting , M., ... Weichselgartner, E. (2015). *Quality Standards for the development, application, and evaluation of measurement instruments in social science survey research.* RatSWD (German Data Forum) Working Papers, 245. Retrieved from http://www.ratswd.de/dl/RatSWD_WP_245.pdf

Sauer, M. (2008). Methodik und Durchführung von standardisierten Befragungen türkeistämmiger Migranten in Deutschland durch die Stiftung Zentrum für Türkeistudien. In K.-S. Rehberg (Ed.), *Die Natur der Gesellschaft: Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel* (pp. 3901-3907). Frankfurt am Main: Campus Verlag.

Schoua-Glusberg, A, Villar, A. (2013). Backtranslation vs. committee approach: An experiment comparing how they perform in questionnaire translation. Presentation given at the FCSM Conference, Washington DC, November 5, 2013.

Schuman, H. (1966). The random probe: A technique for evaluating the validity of closed questions. *American Sociological Review, 31*, 218-222.

Sha, M, Park, H., & Pan, Y. (2012). Translation review and cognitive testing of ACS language assistance guides in multiple languages. Retrieved from http://www.census.gov/library/working-papers/2012/acs/2012_Sha_01.html

Smith, T. W. (2003). Developing comparable questions in cross-national surveys. Harkness, J., van de Vijver, F.J.R., & Mohler, P.P. (Eds.), *Cross-cultural survey methods* (69-91). Hoboken, NJ: Wiley.

Smith, T. (2004). Developing and evaluating cross-national survey instruments. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 431-452). Hoboken, NJ: Wiley.

Snijders T. A. B., & Bosker, R.J. (2011). *Multilevel analysis. An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Stadler, B. (2010). Herausforderungen der Befragung von Migrantinnen und Migranten im Mikrozensus. *Austrian Journal of Statistics, 39*(1&2), 139-145.

Thrasher, J. F., Quah, A. C. K., Dominick, G., Borland R., Driezen, P., Awang, R., et al. (2011). Using cognitive interviewing and behavioral coding to determine measurement equivalence across linguistic and cultural groups: An example from the International Tobacco Control Policy Evaluation Project. *Field Methods, 23*, 439-460.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.

van Deth, J. W. (Eds.) (2013[1998]). *Comparative politics: The problem of equivalence.* ECPR Classics, Colchester: ECPR Press.

van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* Thousand Oaks: Sage.